Data sharing in Radiobiology; towards FAIR

Daniel Berrios[1], Luke T. Slater[2-5], Jack Miller[6], Kristen Peach[7], Sylvain V. Costes[1], and Paul N. Schofield[8]

[1]NASA Ames Research Center, [2]College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, University of Birmingham, UK, [3]Institute of Translational Medicine, University Hospitals Birmingham, NHS Foundation Trust, UK, [4]MRC Health Data Research UK (HDR UK) Midlands, Birmingham , UK, [5]University Hospitals Birmingham NHS Foundation Trust, Edgbaston, Birmingham, UK, [6]KBR, NASA Ames Research Center, Moffett Field, CA, USA, [7]Bionetics, NASA Ames Research Center, [8]Dept of Physiology Development and Neuroscience, University of Cambridge, Downing Street , Cambridge, UK.

Corresponding author: pns12@cam.ac.uk

The value of scientific data depends on their findability, accessibility, integrability and reusability according to the FAIR principles[1]. Together with the sustainability of data preservation and access, these principles underpin the long term benefits of scientific research. Within the domain of radiobiology we have a huge array of data types, themes and complexities which make standardisation of metadata, data structure and data integration very challenging. Moreover, it is clear that, for example, in the area of disaster preparedness, the ready discovery and availability of multiple types of data, for example on biological effects of exposure, climatology, ecology, human behavioural and attitudinal studies, is important for an integrated scientific approach.

Because these data are spread over many databases, journal supplementary information resources and even the computers of the investigators, thier discovery and reuse can be challenging. Despite exhortations from funding agencies and scientific instutions over the past two decades there is still a serious deficit in the willingness and in some cases the ability of investigators to share data, and although much may not be formally „Public domain“, information about the existence of the data, their metadata, and how to obtain them should always be available.

We report the progress of work on three databases, the STORE and the NASA GeneLab and LSDA repositories to leverage the Radiation Biology Ontology (RBO), a structured terminology for metadata that can be used by all radiation biology-relevant databases to unite federated and automated data searches across multiple databases, for example using web services, and through semantic web technologies supporting data discovery.

The initial primary use-cases for RBO were archiving data in the STORE database (https://www.storedb.org/), the repository used for the RadoNorm and Pianoforte Projects among others, and in the NASA Open Science Data Repository (https://osdr.nasa.gov/bio). The scope of radiobiology research ranges from basic physics to radiation oncology to sociolegal studies; no existing ontology had the necessary breadth or depth to fulfil this need. In addition, a formal ontology has the advantage of being usable for machine learning and, importantly, for tasks like data integration, knowledge extraction from the scientific literature and for query extension and data classification. Standardisation of metadata is one of the primary objectives of the FAIR principles for open data; RBO is an important landmark for FAIR-compliant radiation biology data sharing.

The RBO is developed using the open-source tools of GitHub and the OBO Foundry-led Ontology Development Kit, and published through GitHub and the NIH/NCBI BioPortal website. This initial phase of concept modeling has yielded an ontology that has more than 300 declared concepts, with more than 3500 additional concepts imported from other OBO Foundry ontologies with relevance to radiation biology (for example, concepts from the ISO standard Basic Formal Ontology, the Environment Ontology and the Gene Ontology).

We welcome input into the development of RBO and encourage its adoption.

Funding: Euratom research and training program 2014-2018 # 900009, RadoNorm, and U.S. NASA.

[1] Wilkinson, *et al.* (2016). *Scientific data* **3**:160018